

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

APPLICANT/APPELLANT: Norikazu Endo and Benjamin K. Reaves
APPLICATION NO.: 10/661,152
FILING DATE: September 12, 2003
TITLE: Method and System for Adjusting the Voice Prompt of an
Interactive System Based Upon the User's State
EXAMINER: Abdelali Serrou
GROUP ART UNIT: 2626
ATTY. DKT. NO.: 23230-08142

CERTIFICATE OF EFS-WEB TRANSMISSION

Pursuant to 240 OG 45 and the Legal Framework For EFS-Web, I hereby certify that this follow-on correspondence is being officially submitted through the USPTO EFS-Web system from the Pacific Time Zone of the United States on the local date shown below:

Dated: December 4, 2008 By: /Jae Won Song/

Jae Won Song, Reg. No. 59,070

MAIL STOP APPEAL BRIEF- PATENTS
COMMISSIONER FOR PATENTS
P.O. BOX 1450
ALEXANDRIA, VA 22313-1450

REPLY BRIEF

Sir:

This Reply Brief is in reply to the Examiner's Answer mailed on October 31, 2008, and responds to the points made in the Examiner's Answer.

Argument

In the second paragraph on page 8 of the Examiner's Answer regarding claims 1, 15, 28, 40, and 53, the Examiner disagrees with Appellant's argument that Cooper does not disclose or even suggest obtaining utterance parameters that indicate the state of the user by partitioning the utterance into segments and assigning one of a plurality of classifications to **each segment**, where each classification corresponds to at least one of a plurality of states of the user. In this regard, the Examiner's Answer points to statements that appear to be quoted from column 43, line 52 through column, 44, line 6 of Cooper as disclosing this limitation. Specifically, the Examiner's Answer in the second paragraph on page 8 asserts that Cooper discloses a database storing words or phrases that are classified and that the virtual system of Cooper selects the set of classified words depending upon whether the user is, for example, angry or polite, and that such database of Cooper teaches assigning one of a plurality of classifications to each segment of the utterance, as recited in 1, 15, 28, 40, and 53.

Appellants respectfully disagree. This disclosure in column 43, line 52 - column, 44, line 6 of Cooper has nothing to do with obtaining utterance parameters that indicate the state of the user by partitioning the utterance into segments and assigning one of a plurality of classifications to each segment, where each classification corresponds to at least one of a plurality of states of the user. In the inventions of claims 1, 15, 28, 40, and 53, the obtained utterance parameters are eventually used to determine the state of the user (See e.g., claim 1, reciting "determining the state of the user based upon the utterance parameters"). In contrast, this disclosure on "classified words" in column 43, line 52 - column 44, line 6 of Cooper is merely about selecting and using the set of classified words, phrases, or sentences based on the already determined state of the user (angry, polite, etc.). The words or phrases are already

classified (polite discourse, submissive discourse) so that they can be selected to be used by the Cooper virtual system depending upon the already-determined state of the user. However, the “classification” of the words is not used in actually determining the state of the user in Cooper as recited in claims 1, 15, 28, 40, and 53.

In contrast, the inventions of claims 1, 15, 28, 40, and 53 obtain utterance parameters eventually in order to determine the state of the user, by partitioning the utterance into segments and assigning one of a plurality of classifications to each segment, where each classification corresponds to at least one of a plurality of states of the user. The classifications are assigned to each segment of the utterance so that they can be used to eventually determine the state of the user. Such assignment of classifications to each segment for determining the state of the user as recited in claims 1, 15, 28, 40, and 53 is not disclosed or suggested in Cooper.

In the first paragraph on page 9 of the Examiner’s Answer regarding claims 1, 15, 28, 40, and 53, the Examiner asserts that Cooper *inherently* discloses obtaining utterance parameters indicating the state of the user as recited in claims 1, 15, 28, 40, and 53, because otherwise the virtual system would not be able to know the emotional state of the user in order to adapt its behavior accordingly. However, Appellants respectfully disagree, because determining the emotional state of the user does not necessarily mean that utterance parameters have to be obtained as recited in claims 1, 15, 28, 40, and 53, i.e., by assigning one of a plurality of classifications to each segment, where each classification corresponds to at least one of a plurality of states of the user. In fact, Cooper discloses a variety of other methods of determining the state of the user, i.e., based on the user’s voice volume, word-choice, and speech rate (Cooper, column 43, lines 62-65), but none of the examples given in Cooper

correspond to assigning a classification to each segment of the utterance as recited in claims 1, 15, 28, 40, and 53. The very fact that Cooper discloses a variety of different methods of determining the user's state based on the user's utterance, none of which is by assigning one of a plurality of classifications to each segment, is in direct conflict with the Examiner's assertion that Cooper inherently must obtain utterance parameters by assigning a classification to each segment of the utterance, as recited in claims 1, 15, 28, 40, and 53. Inherency requires: "To establish inherency, the extrinsic evidence 'must make clear that the missing descriptive matter is necessarily present in the thing described in the reference, and that it would be so recognized by persons of ordinary skill. Inherency, however, may not be established by probabilities or possibilities. The mere fact that a certain thing may result from a given set of circumstances is not sufficient.' " *In re Robertson*, 169 F.3d 743, 745, 49 USPQ2d 1949, 1950-51 (Fed. Cir. 1999). See MPEP §2112 for the requirements of an anticipation rejection based on inherency. Since Cooper discloses a variety of methods of determining the user's state based on the user's utterance, none of which is by assigning one of a plurality of classifications to each segment, the Cooper method does not have to determine the user's state necessarily by assigning one of a plurality of classifications to each segment as recited in claims 1, 15, 28, 40, and 53.

In the second paragraph on page 9 of the Examiner's Answer regarding claims 1, 15, 28, 40, and 53, it is stated that receiving an utterance and partitioning the utterance into segments is known standard in the speech recognition process. However, the Examiner's Answer still provides no explanation in response to Appellant's argument that Cooper does not disclose that each segment of the utterance is assigned a classification indicating one of a plurality of states of a user. The inventions of claims 1, 15, 28, 40, and 53 require that a

classification is assigned to **each segment** of the utterance, which is still not addressed in the Examiner's Answer.

In the third paragraph on page 9 to the first paragraph on page 10 of the Examiner's Answer regarding claims 1, 15, 28, 40, and 53, it is stated "After an input utterance is received by the virtual assistant system of Cooper, it gets segmented into frames for speech recognition process (column 6, lines 32-37)." However, an exact quote of column 6, lines 32-37 of the Cooper is as follows:

"An Introduction to Virtual Assistant Applications

A VA application allows a user to manage electronic communications and access his or her business's computer resources through a telephone. Using speech recognition and text-to-speech technology, the VA communicates with callers in spoken English."

Note that the above statements from Cooper (column 6, lines 32-37) do not have any disclosure on segmenting of speech into frames, contrary to the Examiner assertions.

Then, in the same third paragraph on page 9 to the first paragraph on page 10 of the Examiner's Answer, it is stated "A frame could be a word, phrase, or sentence. Each frame is approximated by one of a fixed set of possible audio vectors. Those utterance segments are obviously compared with training data. The training data is stored within the database in the form of vectors, wherein each vector refers to a different emotional state; and therefrom determining the emotion state that best matches the data segment, and identifying the emotional state of the user." However, note that the above Examiner's assertions are conclusive statements without any explanation on their basis or rational, and no citation to Cooper is provided to support these assertions. Note that Cooper does not even mention the term

“vector” and the above Examiner’s statements are not disclosed anywhere in Cooper. Rather, these statements appear to be speculation on what *may* occur in the Cooper virtual system. However, it is exactly such speculation that cannot be used as a basis for an anticipation rejection based on inherency. “In relying upon the theory of inherency, the examiner must provide a basis in fact and/or technical reasoning to reasonably support the determination that the allegedly inherent characteristic necessarily flows from the teachings of the applied prior art.” *Ex parte Levy*, 17 USPQ2d 1461, 1464 (Bd. Pat. App. & Inter. 1990). See MPEP 2112 on Requirements of Rejection based on Inherency.

In the second paragraph on page 10 to the first paragraph on page 11 of the Examiner’s Answer regarding claims 3 and 54, it is again conclusively stated that “there is no other way than generating vectors based on the utterance parameters to determine the emotional states of the user, especially when Cooper teaches that the emotional state of the user is based on words spoken by the user.” However, again the Examiner’s Answer still does not offer any rationale or basis as to *why there is no other way* of determining the emotional state of the user other than by generating an utterance parameter vector based upon the utterance parameters and converting the utterance parameter vector to an indication representing the state of the user as recited in claims 3 and 54, which is required for a rejection based on inherency. Such assertion is merely a conclusive statement without any basis. In fact, since Cooper discloses a variety of other methods of determining the state of the user, i.e., based on the user’s voice volume, word-choice, and speech rate (Cooper, column 43, lines 62-65), Cooper itself does disclose *other ways* of determining the emotional state of the user other than by generating an utterance parameter vector based upon the utterance parameters and converting the utterance parameter vector to an indication representing the state of the user as recited in claims 3 and 54.

In the second paragraph on page 11 of the Examiner's Answer regarding claims 3 and 54, the Examiner states, "What has been done by Cooper to determine the emotional state of the user is standard in the field of speech recognition. An utterance is received; the utterance is segmented into segments; each segment is represented by one of a fixed set of possible vector audio parameters, wherein each parameter vector component varies with each different emotional state; and as a result, the emotional state of the user is determined." However, none of these statements in the Examiner's Answer are found in Cooper. Cooper nowhere even mentions representing each segment by one of a fixed set of possible vector audio parameters, wherein each parameter vector component varies with each different emotional state. These statements regarding Cooper are speculation without any basis or support in Cooper. The Examiner's Answer does not provide even a single citation to Cooper to support these statements.

In the third paragraph on page 11 of the Examiner's Answer regarding claims 3 and 54, it is stated, "...using parameters vectors is an inherent step within the process of speech analysis and classification." Such conclusive statement is not permitted for an inherency rejection. The Examiner still does not provide any explanation as to why using parameter vectors is an inherent step in speech analysis and classification, but simply makes a conclusive statement.

In the fourth paragraph on page 11 to the first paragraph on page 12 of the Examiner's Answer regarding claims 4, 17, 30, and 42, the Examiner states, "... Cooper's method of determining vector's probability must be based on the total number of segments within the spoken utterance and the number of segments for each classification, which is a standard method of determining the probability or the likelihood of a word type occurrence." The

Examiner again does not offer any explanation or rationale as to (i) why Cooper's method of determining vector's probability *must be* based on the total number of segments within the spoken utterance and the number of segments for each classification, and (ii) why generating the utterance parameter vector by determining the number of segments for each classification and dividing the number of segments for each classification by a total number of segments in the utterance is such a "standard" method. Again, note that Cooper does not even mention the term "vector" at all anywhere in the document. When Cooper does not even mention the term "vector," it is even more without basis to assume that an utterance parameter vector is necessarily generated at all or necessarily generated in the manner as recited in claims 4, 17, 30, and 42.

In summary, all the claims 1, 3-15, 17-28, 30-40, 42-61 were rejected based on unfounded assertions of inherency without any rationale or explanation as to the reason for such inherency being offered. If the claimed inventions of claims 1, 3-15, 17-28, 30-40, 42-61 were so inherent and "standard" in speech recognition as the Examiner asserts, the Examiner should be able to readily identify and cite references showing why the claimed inventions of claims 1, 3-15, 17-28, 30-40, 42-61 are so inherent and "standard." Nonetheless, such references have not been cited, and unfounded "inherency" rejections are excessively used to reject all the pending claims 1, 3-15, 17-28, 30-40, and 42-61. It is respectfully requested that the Board reverse such erroneous application of the law of inherency and the erroneous rejections of claims 1, 3-15, 17-28, 30-40, and 42-61.

Respectfully submitted,

Dated: December 4, 2008

By: /Jac Won Song/
Jac Won Song, Reg. No. 59,070
Fenwick & West LLP

Silicon Valley Center
801 California Street
Mountain View, CA 94041
Tel.: (650) 335-7164
Fax: (650) 938-5200
Email: jsong@fenwick.com